

패널 승법 계열 시계열 모형의 동질성 검정과 적용¹⁾

이 성 덕²⁾, 김 성 호³⁾, 차 경 엽³⁾

요 약

계절성을 갖는 승법 계열 혼합 시계열 모형들의 동질성 검정을 위하여 Wald 검정 통계량을 구하고 그 극한 분포가 χ^2 -분포함을 보였으며 시뮬레이션 연구를 통하여 뒷받침하였다. 도시 규모가 비슷한 우리 나라 지역별 평균 온도자료를 가지고 이 동질성 검정을 수행하여 시계열을 지역별로 모형화하여 예측한 것과 동질성이 있는 것을 묶고 모형화하여 예측한 것에 대한 예측 오차를 비교하였다.

1. 서론

시계열 모형을 연구함에 있어서 모수 추약의 원칙(principle of parsimony)은 회귀분석에 못지 않게 중요한 원칙이다. 모수의 수가 많으면 모수 추정에 따르는 오차가 커지게 되므로 예측을 정확하게 할 수가 없게 된다. 따라서 모형에 대한 적절한 해석을 위해 가능하며 적은 수의 모수를 갖는 것이 좋다. 만약에 여러 개의 자료들이 동일한 모형에서 왔다고 하는 동질성 가설이 채택된다면, 모형설정에 있어 모수추약을 이룰 수 있고 또한 결합된 자료에 의해 보다 나은 모수 추정치를 얻을 수 있다. 한편 자료들 사이에 유의한 차이가 존재한다 할지라도 표본 크기가 커지면 대표본하에서 어떤 규칙들을 얻을 수 있다.

패널 자료에 대한 동질성 검정을 처음 시도한 것은 Anderson 과 Goodman(1957)의 Markov chain과정에 대한 연구가 그 시초라 할 수 있다. 그 후 Anderson(1978)은 패널 시계열 모형에 대한 동질성 검정에 대하여 연구하였는 바 반복된 측정값들을 가진 자기 회귀 모형에 대하여 최우 추정량을 구하고 동질성 검정에 있어서 우도비 검정 통계량을 사용하여 그 접근 분포가 χ^2 분포함을 보였다. 일반적으로 동질성 검정에는 우도비 통계량, Rao 통계량 그리고 Wald 통계량의 세 검정 통계량이 사용되는데 독립적이고 동일한 분포를 갖는 확률변수들에 대해서는 이 세 검정 통계량들이 동일한 접근 분포를 갖는다는 것이 알려져 왔으나(Serfling, 1980) 시계열은 dependent 확률 변수이므로 일반적인 대수의 법칙이나 중심극한 정리를 이용할 수 없다. 그러나, Basawa, Billard 그리고 Srinivasan(1984)은 세 가지 정규조건하에서 확률변수가 독립적이지 아니더라도, 혹은 동일한 분포를 갖지 않더라도 이 세 검정 통계량이 동일한 극한 분포를 갖는다는 것을 증명했다.

1) 이 논문은 1994도 한국 학술 진흥 재단의 지방대육성 지원과제 연구비에 의해 연구되었음

2) 충북 청주시 흥덕구 개신동 산 48, 충북대학교 자연과학대학 통계학과 부교수, 360-763

3) 충북 청주시 흥덕구 개신동 산 48, 충북대학교 자연과학대학 통계학과, 360-763

이어서 Basawa와 Billard(1989)는 자기상관오차를 가진 회귀모형에 대한 동질성 검정을 연구하였으며 이 때 Wald 통계량을 검정 통계량으로 사용하여 그것의 극한 분포가 χ^2 -분포함을 보였다.

한편 Hwang과 Basawa(1993)는 패널 비선형 자기 회귀 과정에 대한 동질성 검정을 연구했고, 이때 Wald통계량을 검정 통계량으로 사용하여 그 극한 분포를 구하였으며 Billard 와 Meshkani(1989)는 비정상성을 갖는 패널 Markov chain의 동질성 검정 및 대표본 성질을 연구하였다. 본 논문에서는 Box-Jenkins(1976)가 제안한 모형 중 계절성을 갖는 자기회귀모형에 대해 Wald통계량을 이용하여 동질성 검정을 연구하였고 시뮬레이션 연구로 뒷받침하였으며 지역별 우리나라 평균은도자료를 가지고 동질성 검정을 실증 연구하였다.

2. 패널승법계절시계열모형

시계열 $\{X_{t,j}, t=1,2,\dots,n_j, j=1,2,\dots,k\}$ 가 t 시점에서의 j 번째 시계열일 때 이를 패널승법계절시계열 모형이라 하며 패널 SARIMA $(p, d, q) \times (P, D, Q)$ 로 표시하고 다음과 같은 모형식을 갖는다.

$$\phi_{p,j}(B)\Phi_{P,j}(B^{P_s})(1-B)^d(1-B^s)^D X_{t,j} = \theta_{q,j}(B)\Theta_{Q,j}(B^{Q_s})e_{t,j}, \quad (2.1)$$

$$t=1, \dots, n_j, \quad j=1, \dots, k.$$

여기서

$$\begin{aligned} \phi_{p,j}(B) &= 1 - \phi_{1,j}B^1 - \phi_{2,j}B^2 - \dots - \phi_{p,j}B^p, \\ \Phi_{P,j}(B^{P_s}) &= 1 - \Phi_{1,j}B^{1s} - \Phi_{2,j}B^{2s} - \dots - \Phi_{P,j}B^{P_s}, \\ \theta_{q,j}(B) &= 1 - \theta_{1,j}B^1 - \theta_{2,j}B^2 - \dots - \theta_{q,j}B^q, \\ \Theta_{Q,j}(B^{Q_s}) &= 1 - \Theta_{1,j}B^{1s} - \Theta_{2,j}B^{2s} - \dots - \Theta_{Q,j}B^{Q_s}, \end{aligned}$$

그리고 $e_{t,j}$ 는 독립이고 동일한 정규분포에 따르는 $E(e_{t,j})=0$, $V(e_{t,j})=\sigma_e^2$ 인 확률변수이며, B 는 후진연산자이다. 이러한 패널 시계열 모형에 따르는 예는 우리나라의 각 시·도 단위의 월별 경제 지표, 어느 병원에서 여러날 동안 관측한 여러 환자들의 일별 혈압 등 실제적으로 여러 분야에 수많이 존재한다.

3. 패널SAR(1,1) 모형의 통계적 추론

패널승법계절시계열모형에서 $d=D=q=Q=0$ 이고 $p=P=1$ 인 가장 단순한 패널 승법 시계열 모형, SAR(1,1) 모형은 다음과 같다.

$$(1 - \phi_{1,j}B)(1 - \Phi_{1,j}B^s)X_{t,j} = e_{t,j} \quad (3.1)$$

또는

$$X_{t,j} = \phi_{1,j}X_{t-1,j} + \Phi_{1,j}X_{t-s,j} - \phi_{1,j}\Phi_{1,j}X_{t-s-1,j} + e_{t,j}, \quad t=1,2,\dots,n_j, \quad j=1,2,\dots,k$$

각각의 $j=1, \dots, k$ 에 대해 $X_{0,j}=0$ 이라하고 표본으로부터 뽑은 모든 관측치의 수 $n = \sum_{j=1}^k n_j$ 이며, 정상성조건 $|\phi_{1,j}| < 1$, $|\Phi_{1,j}| < 1$ 를 만족할 때 $X_{t,j}$ 는 (3.1)식으로부터 다음과 같이 나타낼 수 있다.

$$\begin{aligned} X_{t,j} &= \{(1-\phi_{1,j}B)(1-\Phi_{1,j}B^s)\}^{-1}e_{t,j} \\ &= \sum_{r=0}^{\infty} \Psi_{r,j} e_{t-r,j}, \end{aligned} \quad (3.2)$$

여기서 $\sum_{r=0}^{\infty} |\Psi_{r,j}| < \infty$ 이다.

(3.1)식과 (3.2)으로부터 계수 $\Psi_{r,j}$ 을 구하면 다음과 같다.

$$\Psi_{r,j} = \phi_{1,j}^h [\phi_{1,j}^{(u+1)s} - \Phi_{1,j}^{(u+1)}] / (\phi_{1,j}^s - \Phi_{1,j}), \quad j=1, \dots, k \quad (3.3)$$

여기서 $r=us+h$, $h=0, 1, \dots, s-1$ 그리고 $u=0, 1, \dots$ 이다.

시계열 $X_{t,j}$ 에 대한 평균과 공분산함수는 다음과 같다.

$$\begin{aligned} E(X_{t,j}) &= \sum_{r=0}^{\infty} \Psi_{r,j} E(e_{t-r,j}) = 0 \\ \gamma_j(h) &= \sigma^2 \sum_{r=0}^{\infty} \Psi_{r,j} \Psi_{r+h,j}, \quad h=0, 1, \dots \end{aligned}$$

실제로 패널 SAR(1,1)모형의 분산을 구하면 다음과 같다.

$$\begin{aligned} V(X_{t,j}) &= \gamma_j(0) = \sigma^2 \sum_{r=0}^{\infty} \Psi_{r,j}^2 \\ &= \sigma^2 (1 + \phi_{1,j}^s \Phi_{1,j}) / \{(1 - \phi_{1,j}^2)(1 - \Phi_{1,j}^2)(1 - \phi_{1,j}^s \Phi_{1,j})\}. \end{aligned}$$

4. 동질성 검정을 위한 통계량과 극한분포

n 개의 표본에 대한 우도함수는 다음과 같이 주어진다.

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^k \sum_{t=1}^{n_j} e_{t,j}^2\right\}$$

여기서 $e_{t,j} = (X_{t,j} - \phi_{1,j}X_{t-1,j} - \Phi_{1,j}X_{t-s,j} + \phi_{1,j}\Phi_{1,j}X_{t-s-1,j})$ 이다.

식(3.1)에 대하여 모수 $\phi_{1,j}$, $\Phi_{1,j}$ 의 최우추정량 $\hat{\phi}_{1,j}$ 와 $\hat{\Phi}_{1,j}$ 은 $\frac{\partial \text{Ln}L}{\partial \phi_{1,j}} = 0$, $\frac{\partial \text{Ln}L}{\partial \Phi_{1,j}} = 0$ 의 해이

며 같은 방법으로 σ^2 의 최우추정량 $\hat{\sigma}^2$ 을 구하면 다음과 같다.

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^k \sum_{t=1}^{n_j} (X_{t,j} - \hat{\phi}_{1,j}X_{t-1,j} - \hat{\Phi}_{1,j}X_{t-s,j} + \hat{\phi}_{1,j}\hat{\Phi}_{1,j}X_{t-s-1,j})^2.$$

모수벡터 $\theta_j^T = (\phi_{1,j}, \Phi_{1,j})$ 라 하고 적당한 초기 추정치 $\theta_{0,j}$ 에서 $e_{t,j}(\theta)$ 를 테일러 급수 전개 (Taylor series expansion)를 하면 다음과 같다.

$$e_{t,j}(\theta_j) \cong e_{t,j}(\theta_{0,j}) - D_{t,j}^T(\theta_j - \theta_{0,j}), \quad j=1, \dots, k, \quad (3.4)$$

여기서 $D_{t,j} = (-\partial e_{t,j}/\partial \phi_{1,j}, -\partial e_{t,j}/\partial \Phi_{1,j})^T$ 인 2×1 벡터이다. 한편 모든 표본에 대한 오차의 합은 다음과 같이 주어지며

$$\sum_{t=1}^{n_j} e_{t,j}^2(\theta_j) \cong \sum_{t=1}^{n_j} \{e_{t,j}(\theta_{0,j}) - D_{t,j}^T(\theta_j - \theta_{0,j})\}^2. \quad (3.5)$$

θ_j 에 대하여 식(3.5)의 우변을 최소로 해주는 $\hat{\theta}_j$ 을 구하면 다음과 같다.

$$\hat{\theta}_j = \theta_{0,j} + \left(\sum_{t=1}^{n_j} D_{t,j} D_{t,j}^T \right)^{-1} \left\{ \sum_{t=1}^{n_j} D_{t,j} e_{t,j}(\theta_{0,j}) \right\}. \quad (3.6)$$

여기서 $\theta_{0,j}$ 는 초기 추정치 $\hat{\theta}_{0,j}$ 에 의해 대체될 필요가 있다. 이때 $\theta_{0,j}$ 에 대체되어진 $\hat{\theta}_{0,j}$ 를 대입하여 얻은 $\hat{\theta}_j$ 를 $\hat{\theta}_j^*$ 라하면 $\hat{\theta}_j^*$ 는 θ_j 의 근사최소제곱추정량이 되고 $\hat{\theta}_j^*$ 는 $\hat{\theta}_j$ 와 같은 극한분포를 가지며 최우추정량 $\hat{\theta}_j = (\hat{\phi}_{1,j}, \hat{\Phi}_{1,j})^T$ 와도 같은 극한분포를 갖는다(Fuller, 1976). 그러므로 식(3.6)로부터 $\hat{\theta}_j$ 의 극한분포를 다음과 같은 형태로부터 구할 수 있다.

$$\sqrt{n_j}(\hat{\theta}_j - \theta_j) = \left(n_j^{-1} \sum_{t=1}^{n_j} D_{t,j} D_{t,j}^T \right)^{-1} \left\{ n_j^{-\frac{1}{2}} \sum_{t=1}^{n_j} D_{t,j} e_{t,j}(\theta_j) \right\}. \quad (3.7)$$

보조정리 4.1. 시계열 $\{X_{t,j}, t=1, \dots, n_j, j=1, \dots, k$ 가 정상성을 만족하는 패널 SAR(1,1) 시계열일 때, 식 (3.7) 우변의 분모는 $n_j \rightarrow \infty$ 일 때,

$$\left(n_j^{-1} \sum_{t=1}^{n_j} D_{t,j} D_{t,j}^T \right) \xrightarrow{p} g_j(\theta_j). \quad (3.8)$$

여기서, $g_j(\theta_j) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

이때, $a_{11} = \gamma_j(0) - 2\Phi_{1,j}\gamma_j(s) + \Phi_{1,j}^2\gamma_j(0)$ 이며, $a_{22} = \gamma_j(0) - 2\phi_{1,j}\gamma_j(1) + \phi_{1,j}^2\gamma_j(0)$, 그리고

$a_{12} = a_{21} = \gamma_j(s-1) - \Phi_{1,j}\gamma_j(1) - \phi_{1,j}\gamma_j(s) + \phi_{1,j}\Phi_{1,j}\gamma_j(0)$ 이며 $\gamma_j(h) = \sigma^2 \sum_{r=1}^{\infty} \Psi_{r,j} \Psi_{r+h,j}$,

$h=0, 1, \dots$ 이고 $\Psi_{r,j}$ 는 식 (3.3)으로부터 얻는다.

보조정리 4.2. 시계열 $\{X_{t,j}, t=1, \dots, n_j, j=1, \dots, k$ 가 정상성을 만족하는 패널 SAR(1,1)시계열이라 할 때 각각의 시계열 $j=1, \dots, k$ 에 대해, $n_j \rightarrow \infty$ 일 때 $n_j^{-\frac{1}{2}} \sum_{t=1}^{n_j} D_{t,j} e_{t,j}(\theta_j)$ 의 극한분포는 다음과 같다.

$$n_j^{-\frac{1}{2}} \sum_{t=1}^{n_j} D_{t,j} e_{t,j}(\theta_j) \xrightarrow{d} N_2(0, \sigma^2 g_j(\theta_j)).$$

보조정리 4.3. 시계열 $\{X_{t,j}, t=1, \dots, n_j$ 이고 $j=1, \dots, k$ 가 정상성을 만족하는 패널 SAR(1,1) 시계열이라고 $\theta_j^T = (\phi_{1,j}, \Phi_{1,j})$ 일 때, 즉 $\theta^T = (\theta_1, \dots, \theta_k)$ 이면 $\sqrt{n}(\hat{\theta} - \theta)$ 의 극한분포는 다음과 같다.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N_{2k}(0, G(\theta)), \quad n \rightarrow \infty \quad (3.9)$$

여기서 $G(\theta) = \sigma^2 \text{Diag}[\{c_1 g_1(\theta_1)\}^{-1}, \dots, \{c_k g_k(\theta_k)\}^{-1}]$ 인 $(2k \times 2k)$ 행렬이 된다.

위의 보조정리 4.1-4.3에 대한 증명은 dependent 확률변수에 대한 대수의 법칙과 m-dependent 과정에 대한 중심극한 정리를 이용한 Lee(1993)을 참고하면 쉽게 구해진다.

패널 SAR(1,1)의 동질성검정을 위한 Wald통계량은 다음과 같이 구해진다. 먼저 패널SAR(1,1) 모형의 동질성 검정을 위한 귀무가설은 다음과 같다.

$$\begin{aligned} H_0 &: \phi_1 = \dots = \phi_k, \quad \Phi_1 = \dots = \Phi_k \\ H_1 &: \text{적어도 하나의 } \phi_j, \Phi_j \text{ 은 같지않다} \end{aligned} \quad (3.10)$$

$\theta_j = (\phi_j, \Phi_j)^T$ 라 놓고 $\eta_j = \theta_j - \theta_1, \quad j=2, \dots, k$ 로 이루어진 벡터 (η_2, \dots, η_k) 로 $(\theta_1, \dots, \theta_k)$ 를 모두 대체시키면 η_j 에 대한 귀무가설은 다음과 같이 된다.

$$\begin{aligned} H_0 &: \eta_j = 0, \quad j=2, \dots, k \\ H_1 &: \text{적어도 하나의 } \eta_j \neq 0 \text{ 이다.} \end{aligned} \quad (3.11)$$

그리고 $\theta = (\theta_1, \dots, \theta_k)^T, \quad \eta = (\eta_2, \dots, \eta_k)^T$ 라 놓으면 패널 SAR(1,1) 모형의 Wald 통계량은 다음과 같다.

$$Q_{1n} = n(\hat{\eta} - \eta)^T J^{-1}(\theta)(\hat{\eta} - \eta) \quad (3.12)$$

여기서,

$$J(\theta) = (\partial \eta / \partial \theta)^T G(\theta) (\partial \eta / \partial \theta)$$

$$(\partial \eta / \partial \theta)^T = \begin{pmatrix} -I & I & 0 & \dots & 0 \\ -I & 0 & I & \dots & 0 \\ -I & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -I & 0 & 0 & \dots & I \end{pmatrix}$$

여기서, $(\partial \eta / \partial \theta)$ 는 $2k \times 2k$ 행렬이고 I는 2×2 단위행렬이다.

정리 4.1. 귀무가설(3.10)하에서 Wald 통계량의 극한 분포는 자유도 $2(k-1)$ 의 χ^2 -분포에 따른다.

[증명] 가설 (3.10)의 귀무가설하에서 (3.9)식은 가설 (3.11)하에서 $n \rightarrow \infty$ 일때 $\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{d} N_{2k-1}(0, J(\eta))$ 로 분포 수렴하며 따라서 Wald 통계량은 동질성 귀무가설하에서 다음과 같은 극한 분포를 갖는다.

$$\begin{aligned} Q_{1n} &= n(\hat{\eta} - \eta)^T J^{-1}(\theta)(\hat{\eta} - \eta) \\ &= n(\hat{\eta})^T J^{-1}(\theta)(\hat{\eta}) \\ &\xrightarrow{d} \chi^2_{(2(k-1))}. \end{aligned}$$

5. 시뮬레이션 연구

패널SAR(1,1) 모형에서 동질성 검정을 위한 Wald 통계량 값을 $\phi_1 = 0.5, \Phi = 0.5$ 인 경우에

패널의 수가 5개, 10개인 경우에 대해 통계패키지 SAS를 이용하여 난수를 발생시켜 시계열 자료를 형성하여 Wald 통계량이 χ^2 분포로 근접하는지를 알아보았다.

표[5-1]은 패널의 수가 5개이고 $\phi_1=0.5, \theta_1=0.5$ 인 SAR(1,1)모형으로부터 시계열 자료 $(X_{1,j}, \dots, X_{n,j}, j=1, \dots, 5, n=1, \dots, 100)$ 을 1000번 추출하여 Wald 통계량의 유의확률을 구한 결과이다.

[표5-1] 패널이 5개인 경험분포들의 빈도수(%)

Mid point	χ^2 with d.f. 8	Wald statistics	Mid point	χ^2 with d.f. 8	Wald statistics
0.5	0.18	0.16	11.5	5.04	5.38
1.5	1.72	1.52	12.5	3.94	4.28
2.5	4.67	4.19	13.5	3.01	3.33
3.5	7.72	7.06	14.5	2.26	2.56
4.5	9.96	9.28	15.5	1.68	1.93
5.5	11.08	10.49	16.5	1.22	1.44
6.5	11.06	10.71	17.5	0.89	1.06
7.5	10.32	10.19	18.5	0.64	0.78
8.5	9.12	9.17	19.5	0.45	0.56
9.5	7.72	7.93	20.5	0.32	0.40
10.5	6.34	6.62	21.5	0.22	0.28

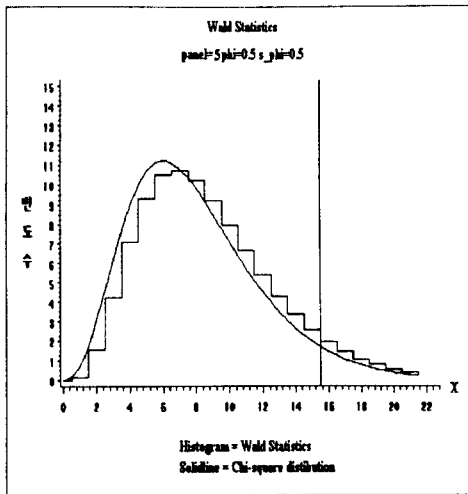
표[5-1]을 보면 Wald 통계량의 유의확률이 χ^2 의 유의확률과 거의 일치함을 알 수 있었다. 패널의 수가 5인 경우 이론적인 극한분포가 χ^2 분포를 따름을 알 수 있었다.

표[5-2]는 패널의 수가 10개이고 $\phi_1=0.5, \theta_1=0.5$ 인 SAR(1,1)모형으로부터 시계열 자료 $(X_{1,j}, \dots, X_{n,j}, j=1, \dots, 10, n=1, \dots, 100)$ 을 1000번 추출하여 Wald 통계량의 유의확률을 구한 결과이다.

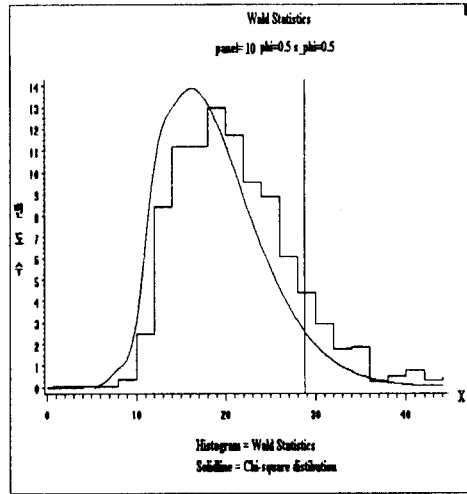
[표5-2] 패널이 10개인 경험분포들의 빈도수(%)

Mid point	χ^2 with d.f. 18	Wald statistics	Mid point	χ^2 with d.f. 18	Wald statistics
0	0	0	24	6.34	6.62
4	0.18	0.16	26	5.04	5.38
6	1.72	1.52	28	3.94	4.28
8	4.67	4.19	30	3.01	3.33
10	7.72	7.06	32	2.26	2.56
12	9.96	9.28	34	1.68	1.93
14	11.08	10.49	36	1.22	1.44
16	11.06	10.71	38	0.89	1.06
18	10.32	10.19	40	0.64	0.78
20	9.12	9.17	42	0.45	0.56
22	7.72	7.93	44	0.32	0.40

표[5-2]을 보면 Wald 통계량의 유의확률이 χ^2 유의확률보다 약간 크게 나왔다. 즉 동질성 검정에 있어 동질성이 없는 시계열 모형을 동질성이 있다고 판단할 확률이 조금 커졌음을 의미한다. 그림[5-1]과 [5-2]는 위 결과들을 그림으로 나타낸 것으로 Wald통계량의 점근 분포가 χ^2 분포에 근사한 것을 나타낸다.



[그림5-1] 패널이 5개일 때
Wald 통계량의 분포도



[그림5-2] 패널이 10개일 때
Wald 통계량의 분포도

6. 사례연구

본 논문에서 사용된 자료는 1971년부터 1992년까지의 우리나라 지역별 평균기온에 대한 자료이다. 자료 168개 중 잔차분석을 위해 24개의 자료를 제외시킨 144개의 자료를 이용하여 패널계절자기회귀모형의 동질성검정을 실시하였다. 지역들의 평균기온은 모형식별 →모수추정 →추정된 모형의 적합성을 진단하는 단계를 거쳐 최종적으로 SAR(1,1) 모형을 따름을 알 수 있었다. 여러 지역(충무, 전주, 청주, 군산, 서산, 수원)들간의 시계열 모형을 수립하고 지역들간의 유의한 차이가 있는지 없는지 동질성 검정을 하기 위하여 Wald 통계량을 사용하였다. 위 시계열 자료는 다음과 같은 패널 SAR(1,1)모형을 갖는다.

$$X_{t,j} = \phi_{1,j} X_{t-1,j} + \phi_{1,j} X_{t-12,j} + \phi_{1,j} \phi_{1,j} X_{t-13} + e_{t,j}, \quad e_{t,j} \text{는 } N(0,1).$$

여기서, j=1은 충무, j=2는 전주, j=3은 청주, j=4는 군산, j=5는 서산, j=6은 수원이며, 각각의 모수 추정 값은 [표6-1]과 같다. 지역들간의 동질성 검정을 위한 귀무가설은 다음과 같아진다.

$$H_0 : \phi_{1,j} = \phi_{1,1}, \quad \phi_{1,j} = \phi_{1,1}, \quad j=1, \dots, 6$$

$$H_1 : \text{적어도 하나의 } \phi_{1,j} \neq \phi_{1,1}, \quad \phi_{2,j} \neq \phi_{1,1}, \quad j=1, \dots, 6.$$

[표6-2]는 각 지역(충무, 전주, 청주, 군산, 서산, 수원)들의 평균온도를 각각의 계절자기회귀모형으로 예측한 값과 귀무가설하에서 pooling한 패널계절자기회귀모형으로 예측한 값을 실제값과 비교하여 예측오차를 구한 표이다. 즉 $\hat{e}_i(t) = X_{t+1} - \hat{X}_i(t)$ 을 구하여 $\hat{e}_i(t)$ 의 평균과 분산을 비교한 것이다. 여기서 X_{t+1} 은 $t+1$ 시점에서의 실제값이며 $\hat{X}_i(t)$ 은 t 시차 후의 예측값으로 그 차이가 예측오차(prediction error)가 된다.

[표6-1] 지역들의 최우추정량

	충무	전주	청주	군산	서산	수원
$\hat{\phi}_1$	0.33154	0.30085	0.29164	0.32992	0.33187	0.26798
$\hat{\phi}_1$	-0.61979	-0.61235	-0.55338	-0.55851	-0.56524	-0.55803

SAS/IML을 이용하여 구한 Wald 통계량 $Q_{1n} = n(\hat{\eta} - \eta)J^{-1}(\theta)(\hat{\eta} - \eta)$ 를 구한 결과는 다음과 같다.

$$Q_{1n} = 1.55516025$$

이 값은 유의수준 $\alpha = 0.05$ 에서 자유도 $2(k-1)$ 을 따르는 χ^2 분포의 임계치보다 작으므로 귀무가설을 채택하게 된다. 그러므로 각 지역들간에는 유의한 차가 없다. 즉 동질성을 만족하는 자료이다. 동질성 귀무가설하에서 ϕ_H 와 ϕ_H 의 최우추정량은 $\hat{\phi}_{1,H} = 0.28138$, $\hat{\phi}_{1,H} = -0.5001$ 이다.

[표6-2] 지역별 모형과 패널 모형의 예측오차비교

		충무	전주	청주	군산	서산	수원
지역별모형	$E(\hat{e}_i)$	-0.6563	-0.7349	-0.6497	-0.6415	-0.5197	-0.6246
	$V(\hat{e}_i)$	0.82394	1.0984	1.2502	0.9175	0.8801	1.2712
패널모형	$E(\hat{e}_i)$	-0.5157	-0.6024	-0.5544	-0.55171	-0.4293	-0.5137
	$V(\hat{e}_i)$	0.7520	1.0470	1.0281	0.8957	0.8535	1.2323

[표6-2]를 보면 각지역 예측오차들의 평균과 분산의 값들이 각 지역별 계절자기회귀모형으로 예측한 값보다 패널모형으로 pooling한 모형으로 예측하였을 때 더 좋은 결과를 얻을 수 있음을 알 수 있다. 즉 각 지역별 시계열 모형에 있어서의 예측오차 및 그 분산이 동질성을 만족한 패널 모형의 그것보다 큰 것을 알 수 있다. 우리나라의 각지역의 평균기온은 동질성을 만족하는 시계열이며 패널모형으로 자료를 처리하는 것이 바람직함을 알 수 있다.

7. 결 론

패널 SAR(1,1) 모형의 동질성 검정을 위한 Wald 통계량의 이론적인 극한분포가 χ^2 분포로 근접하는지에 대하여 시뮬레이션을 통해 유의확률을 구해 연구한 결과와 SAR(1,1)을 따르는 우리나라

라 지역별 평균온도를 가지고 동질성검정을 통해 하나의 패널 모형으로 분석한 결과는 다음과 같다.

첫째, 동질성 검정을 위한 Wald 통계량의 유의확률은 패널의 수가 5개일때는 χ^2 분포의 유의확률과 거의 일치함을 알 수 있었다. 그러나 패널의 수가 10개로 증가하였을 때에는 Wald 통계량의 유의확률이 χ^2 분포의 유의확률보다 약간 큼을 알 수 있었다.

둘째, 우리 나라 지역별 평균온도의 자료를 가지고 동질성 검정을 실시한 결과 지역별 자료는 동질성을 만족하는 시계열이었다. 그리고 각 지역별계절자기회귀모형으로 미래시점에 대해 예측하였을 때보다 패널계절자기회귀모형으로 묶어서 미래시점에 대해 예측을 한 것이 예측오차가 작음을 알 수 있었다. 그러므로 동질성을 만족하는 시계열은 패널모형으로 자료를 처리하는 것이 모수 추약의 원칙에 충실할뿐만 아니라 보다 좋은 예측값을 얻을 수 있다. SAR(p, P)모형으로의 확장 과 시뮬레이션에 있어서의 패널수의 확장은 더 연구해야할 과제라 생각된다.

참고문헌

- [1] Anderson. T. W.(1978). Repeated Measurement on Autoregressive Processes, *Journal of the American Statistical Association*, 73, 371-378.
- [2] Anderson. T.W. and Goodman, L.A. (1957). Statistical inference about Markov Chains *Annals of Mathematical Statistics*, 28, 89-110.
- [3] Basawa, I.V. , Billard, L. (1989). Large sample inference for a regression model with autocorrelated errors, *Biometrika*, 76, 283-288.
- [4] Basawa. I. V., Billard. L. and Srinivasan. R.(1984). Large Sample Tests of Homogeneity for Time Series Models, *Biometrika*, 71, 203-206.
- [5] Billard, L. and Meshkani, M.R.(1989). Nonstationary Markov Chain-Panel study, *Technical Report No.117*, University of Georgia.
- [6] Box. G. E. P. and Jenkins. G. M.(1976). *Time Series Analysis : Forecasting and Control* 2nd ed., Holden Day Inc., California.
- [7] Fuller. W. A.(1976). *Introduction to Statistical Time Series*, John Wiley, NewYork.
- [8] Hwang,S.Y. and Basawa, I.V. (1993). Large sample inference based on multiple observation from nonlinear autoregressive processes, *Technical Report No.92-17*, University of Georgia.
- [9] Lee.S.D.(1993). Test of Homogeneity for a Panel of Seasonal Autoregressive Processes, *Journal of the Korean Statistical Society*, Vol. 22, No. 1, 125-132.
- [10] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley and Song, New york